第27卷 第5期 工 程 数 学 学 报 Vol. 27 No. 5
2010年10月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Oct. 2010

文章编号:1005-3085(2010)05-0771-10

响应变量随机缺失下的变系数部分线性模型的经验似然推断*

赵培信[1]， 薛留根[2]

(1- 河池学院数学系，广西 宜州 546300； 2- 北京工业大学应用数理学院，北京 100124)

摘 要: 本文考虑了响应变量随机缺失下的变系数部分线性模型的估计问题。利用经验似然方法，给出了参数部分的调整经验似然比函数，证明其渐近服从标准卡方分布。进而构造了参数部分的置信域，得到了其极大经验似然估计的最优参数收敛速度和渐近半参数有效界。模拟结果表明调整经验似然方法优于未调整的经验似然方法。

关键词: 变系数部分线性模型；经验似然；置信域；缺失数据

分类号: **AMS**(2000) 62G05; 62G20 **中图分类号**: O212.7 **文献标识码**: A

1 引言

设有 n 个个体的样本，(Y_i, X_i, Z_i, U_i)，$i = 1, \cdots, n$，本文考虑如下的变系数部分线性模型

$$Y_i = X_i^T\theta(U_i) + Z_i^T\beta + \epsilon_i, \tag{1}$$

其中 $\beta = (\beta_1, \cdots, \beta_q)^T$ 为 q 维未知向量，$\theta(\cdot) = (\theta_1(\cdot), \cdots, \theta_p(\cdot))^T$ 为 p 维未知向量函数。Y_i 为响应变量，X_i 为 p 维协变量，Z_i 为 q 维协变量，U_i 为一维协变量。不失一般性，设 U_i 在区间 $[0,1]$ 上取值。并且假定 $E(\epsilon_i \,|\, U_i, X_i, Z_i) = 0$。

模型 (1) 包含了几种被广泛研究的模型，例如线性模型[1,2]，部分线性模型[3-6] 以及变系数模型[7,8] 等都是模型 (1) 的特殊情况。对模型 (1)，在数据不含有缺失的情况下已有大量的文献进行研究[9-11]，但是在实际应用中，响应变量往往因某种原因而产生缺失。当响应变量含有缺失时，现有的统计推断方法将不能直接应用。如果仅仅用可以完全观测到的样本进行统计推断，得到的估计往往会产生偏差，而且一般不是渐近有效的。关于缺失数据问题的研究，目前已有大量的文献对缺失数据下的线性模型[1,2] 以及部分线性模型[12,13] 进行了研究。但是关于缺失数据下的变系数部分线性模型的估计问题，目前还没有相关的文献进行研究。

本文在响应变量随机缺失下，利用经验似然方法对模型 (1) 参数部分的估计问题进行研究。给出了参数部分的调整经验似然比函数，并证其渐近服从标准卡方分布，进而构造了 β 的置信域。另外本文还证明了参数部分的极大经验似然估计达到了最优的参数收敛速度并且是渐近有效的。最后通过数据模拟，对调整的经验似然方法与未调整的经验似然方法的有限样本性质进行了比较。模拟结果显示本文的方法优于未调整的经验似然方法。

收稿日期: 2008-10-29. **作者简介**: 赵培信 (1981年7月生)，男，博士，讲师. 研究方向：非参数统计.

*__基金项目__: 国家自然科学基金 (10871013)；北京市自然科学基金 (1102008)；广西自然科学基金 (2010GXNSFB-013051)；河池学院研究生科研启动基金 (2008QS-N014).

2 参数部分的调整经验似然推断

设 $(Y_i, X_i, Z_i, \delta_i, U_i)$, $i = 1, \cdots, n$，为来自模型 (1) 的一个不完全随机样本，其中 X_i, Z_i 和 U_i 可以完全观测，当 $\delta_i = 1$ 时，Y_i 可以观测，$\delta_i = 0$ 时，Y_i 缺失。本文假定 Y 为随机缺失，即

$$P\big(\delta = 1 \,|\, Y, X, Z, U\big) = P\big(\delta = 1 \,|\, X, Z, U\big).$$

由模型 (1) 可知

$$\delta_i Y_i = \delta_i X_i^T \theta(U_i) + \delta_i Z_i^T \beta + e_i, \tag{2}$$

其中 $e_i = \delta_i \epsilon_i$ 为独立同分布的随机变量，并且 $E(e_i \,|\, X_i, Z_i, U_i) = 0$。模型 (2) 为基于完全观测数据的变系数部分线性模型，因此利用文献 [9] 的方法，对给定的 β，最小化 (3) 式则得到 $\theta(u)$ 的局部最小二乘估计 $\hat{\theta}(u)$。

$$\sum_{i=1}^{n} \Big\{ Y_i - Z_i^T \beta - \sum_{k=1}^{p} \big[a_k + b_k(U_i - u)\big] X_{ik} \Big\}^2 \delta_i K_h(u - U_i), \tag{3}$$

其中 $K_h(\cdot) = h^{-1} K(\cdot/h)$，$K(\cdot)$ 为核函数，h 为带宽，X_{ik} 为 X_i 的第 k 个元素。记

$$Z = \big(Z_1, \cdots, Z_n\big)^T, \quad Y = \big(Y_1, \cdots, Y_n\big)^T,$$

I_p 为 p 阶单位阵，0_p 为 p 阶 0 矩阵

$$S(u) = (I_p, 0_p)(D_u^T \Omega_u D_u)^{-1} D_u^T \Omega_u \equiv \big(S_1(u), \cdots, S_n(u)\big),$$

其中

$$\Omega_u = \mathrm{diag}\big(\delta_1 K_h(u - U_1), \cdots, \delta_n K_h(u - U_n)\big)$$

为 $n \times n$ 对角阵，并且

$$D_u = \begin{pmatrix} X_1 & \cdots & X_n \\ h^{-1}(U_1 - u)X_1 & \cdots & h^{-1}(U_n - u)X_n \end{pmatrix}^T$$

为 $n \times 2p$ 的矩阵，则有

$$\hat{\theta}(U_i) = \sum_{k=1}^{n} S_k(U_i)(Y_k - Z_k^T \beta). \tag{4}$$

把 (4) 式代入 (1) 式，并简单计算得

$$Y_i - X_i^T \hat{g}(U_i) \approx \big[Z_i - \hat{\mu}(U_i)^T X_i \big]^T \beta + \epsilon_i, \tag{5}$$

其中

$$\hat{\mu}(u) = \sum_{k=1}^{n} S_k(u) Z_k^T, \quad \hat{g}(u) = \sum_{k=1}^{n} S_k(u) Y_k.$$

引入辅助随机向量

$$\hat{\eta}_i(\beta) = \frac{\delta_i}{\sqrt{\hat{\Delta}(U_i)}} \big(Z_i - \hat{\mu}(U_i)^T X_i\big) \big\{ Y_i - X_i^T \hat{g}(U_i) - (Z_i - \hat{\mu}(U_i)^T X_i)^T \beta \big\}, \tag{6}$$

其中调整因子 $\hat{\Delta}(u)$ 为 $\Delta(u)=P(\delta_i=1\,|\,U_i=u)$ 的核估计，即

$$\hat{\Delta}(u)=\sum_{i=1}^{n}\omega_{ni}(u)\delta_i,$$

这里

$$\omega_{ni}(u)=\frac{K\big((u-U_i)/h\big)}{\sum_{j=1}^{n}K\big((u-U_j)/h\big)}.$$

由 (5) 式可得 $E\{\hat{\eta}_i(\beta)\}=o(1)$。根据这一信息，$\beta$ 的经验似然比函数可定义为

$$\hat{R}(\beta)=-2\max\Big\{\sum_{i=1}^{n}\log(np_i)\,\Big|\,p_i\geq 0,\ \sum_{i=1}^{n}p_i=1,\ \sum_{i=1}^{n}p_i\hat{\eta}_i(\beta)=0\Big\},$$

其中 $p_i=p_i(\beta)$, $i=1,\cdots,n$。对任意给定的 β，假设 0 在点 $(\hat{\eta}_1(\beta),\cdots,\hat{\eta}_n(\beta))$ 所构成的凸集内部，则 $\hat{R}(\beta)$ 存在唯一的解。利用 Lagrange 乘子法，可以把 $\hat{R}(\beta)$ 表示为

$$\hat{R}(\beta)=2\sum_{i=1}^{n}\log\big(1+\lambda^T\hat{\eta}_i(\beta)\big), \tag{7}$$

其中 $\lambda=\lambda(\beta)$ 为 q 维向量，且满足

$$\sum_{i=1}^{n}\frac{\hat{\eta}_i(\beta)}{1+\lambda^T\hat{\eta}_i(\beta)}=0. \tag{8}$$

定理 1 设第 4 节中的条件 $C1$-$C5$ 成立，如果 β 为参数真值，那么

$$\hat{R}(\beta)\stackrel{\mathcal{L}}{\longrightarrow}\chi_q^2,$$

其中 $\stackrel{\mathcal{L}}{\longrightarrow}$ 表示依分布收敛，χ_q^2 为自由度为 q 的卡方分布。

以 $\chi_q^2(1-\alpha)$ 记 χ_q^2 的 $1-\alpha$ 分位数，$0<\alpha<1$。由定理 1 可以得到 β 的近似 $1-\alpha$ 置信域为

$$C_\alpha(\beta)=\big\{\,\beta\,|\,\hat{R}(\beta)\leq\chi_q^2(1-\alpha)\,\big\}.$$

通过最小化 (7) 式可以得到 β 的一个估计 $\hat{\beta}$，称之为极大经验似然估计。记

$$\hat{\Gamma}=\frac{1}{n}\sum_{i=1}^{n}\big(Z_i-\hat{\mu}(U_i)^TX_i\big)\big(Z_i-\hat{\mu}(U_i)^TX_i\big)^T,$$

如果矩阵 $\hat{\Gamma}$ 是可逆的，那么利用 Qin 和 Lawless[14] 的定理 1 的类似证明可知，β 的极大经验似然估计可以表示为

$$\hat{\beta}=\hat{\Gamma}^{-1}\frac{1}{n}\sum_{i=1}^{n}\big(Z_i-\hat{\mu}(U_i)^TX_i\big)\big(Y_i-X_i^T\hat{g}(U_i)\big)+o_p(n^{-\frac{1}{2}}). \tag{9}$$

令

$$\Psi(\cdot)=E\big(X_iX_i^T\,|\,U_i=\cdot\big),\quad \Phi(\cdot)=E\big(X_iZ_i^T\,|\,U_i=\cdot\big),\quad \sigma^2(\cdot)=E\big(\epsilon_i^2\,|\,U_i=\cdot\big),$$

下面的定理给出了 $\hat{\beta}$ 的渐近正态性。

定理2 设第4节中的条件$C1$-$C5$成立，那么

$$\sqrt{n}\,(\hat{\beta}-\beta)\xrightarrow{\mathcal{L}}N(0,\Sigma),$$

其中

$$\Sigma=E\big\{\,\sigma^2(U)\,[\,E(ZZ^T)-\Phi(U)^T\Psi(U)^{-1}\Phi(U)\,]^{-1}\,\big\}.$$

由定理2知$\hat{\beta}=\beta+O_p(n^{-\frac{1}{2}})$，即$\hat{\beta}$达到了最优的参数收敛速度。另外，由文献[15]可知Σ为半参数有效信息界。因此，本文提出的调整经验似然方法所得的估计是半参数有效的。

3 数字模拟分析

为实施模拟，我们从模型$Y=\sin(0.5\pi U)X+1.5Z+\epsilon$中产生数据，其中$X\sim N(0,1)$，$U$和$Z$均为$[-2,2]$上的均匀分布。$Y$由模型产生，其中模型误差$\epsilon$服从均值为0，方差为0.5的正态分布。在模拟过程中，取$n=100$，实验重复1000次。并且缺失概率$\Delta(u)$分别取如下三种情况来代表响应变量的不同缺失水平：

1) 如果$|U|\le 0.5$，则$\Delta(u)=0.9+0.2U$，否则$\Delta(u)=0.95$；

2) 如果$|U|\le 0.5$，则$\Delta(u)=0.7+0.2U$，否则$\Delta(u)=0.75$；

3) 如果$|U|\le 0.5$，则$\Delta(u)=0.4+0.2U$，否则$\Delta(u)=0.45$。

这三种情况对应的数据Y的缺失概率分别平均为10%，30%和60%。核函数取$K(u)=0.75(1-u^2)_+$，用“去一个体”交叉证实法选取带宽h_{CV}，使其满足

$$CV(h)=\frac{1}{n}\sum_{i=1}^{n}\delta_i\big\{\,Y_i-X_i^T\hat{\theta}_{-i}(U_i)-Z_i^T\hat{\beta}_{-i}\,\big\}^2$$

达到最小，其中$\hat{\theta}_{-i}(U_i)$和$\hat{\beta}_{-i}$分别为去掉第i个观测值后$\theta(U_i)$和β的估计量，$\hat{\theta}_{-i}(U_i)$可通过把$\hat{\beta}_{-i}$代入(4)式得到。

对于参数β，我们对本文提出的调整经验似然(ADEL)方法与未调整经验似然方法(NAEL)进行比较。后者是只用完全观测到的样本，直接把(6)式中的调整因子$\hat{\Delta}(U_i)^{-\frac{1}{2}}\delta_i$用$\delta_i$所代替，来构造经验似然比函数。我们分别计算了1000次模拟的绝对偏差$|\hat{\beta}-\beta|$的平均值(Bias)，标准误差的平均值(SE)，置信水平为95%的置信区间长度的平均值(Len)和对应的覆盖概率(Cov)的平均值。模拟结果见表1。

表1: 响应变量Y在三种缺失情况下的模拟结果

	ADEL				NAEL			
$1-E(\Delta)$	Bias	SE	Len	Cov	Bias	SE	Len	Cov
10%	0.055	0.041	0.221	0.954	0.054	0.037	0.223	0.958
30%	0.071	0.037	0.261	0.945	0.169	0.029	0.259	0.640
60%	0.176	0.027	0.363	0.907	0.390	0.017	0.362	0.217

从表 1，我们可以得到如下结论：

(i)　随着缺失概率的增大，利用 ADEL 方法和利用 NAEL 方法所得到的绝对偏差 (Bias) 都逐渐增大。但利用 ADEL 方法所得偏差的增大速度明显小于利用 NAEL 方法所得偏差的增大速度。

(ii)　ADEL 方法和 NAEL 方法均能给出较小的标准误差 (SE)，并且两种方法所得的标准误差对缺失概率的变化都不是太敏感。另外，NAEL 方法给出的标准误差略微小于 ADEL 方法给出的标准误差，这主要是由于 ADEL 方法中的调整因子含有估计量缘故。

(iii)　尽管利用 ADEL 方法和 NAEL 方法所得估计的置信区间长度相差不大，但当缺失概率较大时，NAEL 估计具有较大的绝对偏差，从而导致其覆盖概率是非常不准确的。

总之，从模拟结果来看本文提出的 ADEL 方法具有优良的有限样本性质。并且当缺失概率较大时，ADEL 方法明显优于 NAEL 方法。

4　定理的证明

为书写方便，下文用 c 表示正常数，每次出现时可以代表不同的值。在证明本文的主要结果之前，首先给出一些正则化条件。

$C1$:　带宽满足 $h=Cn^{-\frac{1}{5}}$，其中 $C>0$ 为某给定常数。

$C2$:　核函数 $K(\cdot)$ 是对称的概率核函数，且有

$$\int u^4K(u)\mathrm{d}u<\infty.$$

$C3$:　对任给的 $u\in(0,1)$，$f(u)$, $\Phi(u)$, $\Psi(u)$, $\sigma^2(u)$ 以及 $\theta(u)$ 在 u 点均二次连续可微，其中 $f(\cdot)$ 为 U 的密度函数。

$C4$:

$$\sup_{0\le u\le 1}E\big(\epsilon_i^4\,|\,U_i=u\big)<\infty,\qquad \sup_{0\le u\le 1}E\big(X_{ir}^4\,|\,U_i=u\big)<\infty,$$

且关于 u 连续，$i=1,\cdots,n$, $r=1,\cdots,p$，其中 X_{ir} 是 X_i 的第 r 个分量。

$C5$:　对给定的 u，$\Psi(u)$ 是正定矩阵。

下面首先给出几个引理。

引理 1　设 $a_1,a_2,\cdots,a_n$；$b_1,b_2,\cdots,b_n\,(b_1\ge b_2\ge\cdots\ge b_n)$ 为两实数序列。令

$$S_k=\sum_{i=1}^k a_i,$$

则有

$$\Big|\sum_{i=1}^n a_ib_i\Big|\le c\max_{1\le i\le n}|b_i|\max_{1\le i\le n}|S_i|.$$

证明　利用 Abel 不等式，经简单计算可得到本引理的证明。

引理 2　设 e_i, $i=1,\cdots,n$，为一相互独立的随机变量序列，且满足 $E(e_i)=0$ 和 $E(e_i^2)<c<\infty$，则有

$$\max_{1\le k\le n}\Big|\sum_{i=1}^k e_i\Big|=O_p\big(\sqrt{n}\log n\big).$$

进一步，令 $(j_1,j_2,\cdots,j_n)$ 为 $(1,2,\cdots,n)$ 的任意置换，那么有

$$\max_{1\le k\le n}\Big|\sum_{i=1}^{k}e_{j_i}\Big|=O_p\big(\sqrt{n}\log n\big).$$

证明 利用 Kolmogrov 不等式，我们可以证明本引理。

引理3 设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 为一独立同分布的随机向量，其中 Y_i 是一维随机变量，并且

$$E|Y_1|^s<\infty,\quad \sup_x\int|y|^sf(x,y)\mathrm{d}y<\infty,$$

这里 $f(\cdot,\cdot)$ 是 (X,Y) 的联合密度函数。设 $K(\cdot)$ 为一有界支撑上的非负函数，并满足 Lipschitz 条件，那么有

$$\sup_x\Big|\frac{1}{n}\sum_{i=1}^{n}\big\{K_h(X_i-x)Y_i-E\,[\,K_h(X_i-x)Y_i\,]\big\}\Big|=O_p\Big(\Big\{\frac{\log(1/h)}{nh}\Big\}^{\frac12}\Big),$$

其中 $n^{2\delta-1}h\longrightarrow\infty,\ \delta<1-s^{-1}$。

证明 证明见 Mack 和 Silverman[16]。

引理4 设条件 $C1$-$C5$ 成立，则有

$$\sup_{0<u<1}\big\|\hat\mu(u)-\Psi^{-1}(u)\Phi(u)\big\|=O_p(C_n),$$

$$\sup_{0<u<1}\big\|\hat g(u)-\Psi^{-1}(u)\Phi(u)\beta-\theta(u)\big\|=O_p(C_n),$$

其中 $\hat\mu(u)$ 和 $\hat g(u)$ 由 (5) 式所定义，并且

$$C_n=h^2+\Big(\frac{\log(1/h)}{nh}\Big)^{\frac12}.$$

证明 结合引理3，利用类似文献 [9] 中 (7.1) 式的证法可知，对 $u\in(0,1)$，一致有

$$D_u^T\Omega_uZ=n\Delta(u)f(u)\Phi(u)\otimes(1,0)^T\big\{1+O_p(C_n)\big\},\tag{10}$$

$$D_u^T\Omega_uD_u=n\Delta(u)f(u)\Psi(u)\otimes\begin{pmatrix}1&0\\0&\int s^2K(s)\mathrm{d}s\end{pmatrix}\big\{1+O_p(C_n)\big\},\tag{11}$$

其中 $\Delta(u)=P(\delta_i=1\,|\,U_i=u)$。再结合 $\hat\mu(u)$ 的定义，经简单计算可得

$$\hat\mu(u)=(I_p,0_p)(D_u^T\Omega_uD_u)^{-1}D_u^T\Omega_uZ=\Psi^{-1}(u)\Phi(u)\big\{1+O_p(C_n)\big\},$$

对 $u\in(0,1)$ 一致成立。利用类似的方法可以证明第二式，这就完成了引理4的证明。

引理5 设条件 $C1$-$C5$ 成立，如果 β 是参数真值，则

$$\frac{1}{\sqrt n}\sum_{i=1}^{n}\hat\eta_i(\beta)\stackrel{\mathcal{L}}{\longrightarrow}N(0,B),$$

其中 $B=E\big\{\sigma^2(U)\,[\,E(ZZ^T)-\Phi^T(U)\Psi^{-1}(U)\Phi(U)\,]\big\}$。

证明　结合 (7) 式，经简单计算可得

$$
\begin{aligned}
\hat{\eta}_i(\beta) \;=\;& \frac{\delta_i}{\sqrt{\hat{\Delta}(U_i)}}\big[Z_i-\mu(U_i)^TX_i\big]\epsilon_i+\frac{\delta_i}{\sqrt{\hat{\Delta}(U_i)}}\big[\mu(U_i)-\hat{\mu}(U_i)\big]^TX_i\epsilon_i\\
&+\frac{\delta_i}{\sqrt{\hat{\Delta}(U_i)}}\big[Z_i-\mu(U_i)^TX_i\big]X_i^T\big[\theta(U_i)-\hat{g}(U_i)+\hat{\mu}(U_i)\beta\big]\\
&+\frac{\delta_i}{\sqrt{\hat{\Delta}(U_i)}}\big[\mu(U_i)-\hat{\mu}(U_i)\big]^TX_iX_i^T\big[\theta(U_i)-\hat{g}(U_i)+\hat{\mu}(U_i)\beta\big]\\
\equiv\;& J_{i1}+J_{i2}+J_{i3}+J_{i4}.
\end{aligned}
$$

注意到在条件 $C1, C2$ 下有

$$
\sup_u|\hat{\Delta}(u)-\Delta(u)|=O_p\big((nh)^{-\frac{1}{2}}\big).
$$

再利用对 $\frac{1}{\sqrt{\hat{\Delta}(u)}}$ 在 $\Delta(u)$ 点进行 Taylor 展开，并经简单计算可得

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}J_{i1} \;=\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\sqrt{\Delta(U_i)}}\big[Z_i-\mu(U_i)^TX_i\big]\epsilon_i\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\Delta(U_i)^{3/2}}\big(\hat{\Delta}(U_i)-\Delta(U_i)\big)\big[Z_i-\mu(U_i)^TX_i\big]\epsilon_i+o_p(1)\\
\equiv\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_{i1}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_{i2}+o_p(1).
\end{aligned}
\tag{12}
$$

注意到 $E(I_{i1})=0$, $\mathrm{Var}(I_{i1})=B+o(1)$，利用中心极限定理可知 $n^{-\frac{1}{2}}\sum\limits_{i=1}^{n}I_{i1}\xrightarrow{\mathcal{L}}N(0,B)$。下证 $n^{-\frac{1}{2}}\sum\limits_{i=1}^{n}I_{i2}\xrightarrow{p}0$。令

$$
b(U_i)=\hat{\Delta}(U_i)-\Delta(U_i),\quad a_i=\frac{\delta_i}{\Delta(U_i)^{3/2}}\big(Z_i-\mu(U_i)^TX_i\big)\epsilon_i
$$

为 $q\times1$ 维向量。另外，令 $a_{i,s}$ 为 a_i 的第 s 个分量，$(b(U_{t_i}), i=1,\cdots,n)$ 为 $(b(U_i), i=1,\cdots,n)$ 的某一置换，使得 $b(U_{t_1})\ge b(U_{t_2})\ge\cdots\ge b(U_{t_n})$，对应的 $(a_{i,s}, i=1,\cdots,n)$ 记为 $(a_{t_i,s}, i=1,\cdots,n)$。设 $I_{i2,s}$ 为 I_{i2} 的第 s 个分量，由引理 1 至引理 4 可得

$$
\begin{aligned}
\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_{i2,s}\Big| \;=\;& \frac{1}{\sqrt{n}}\Big|\sum_{i=1}^{n}b(U_{t_i})a_{t_i,s}\Big|\le\frac{1}{\sqrt{n}}\sup_{0\le u\le1}|b(u)|\max_{1\le k\le n}\Big|\sum_{i=1}^{k}a_{t_i,s}\Big|\\
=\;& \frac{1}{\sqrt{n}}O_p(C_n)O_p(\sqrt{n}\log n)=o_p(1),
\end{aligned}
$$

即 $n^{-\frac{1}{2}}\sum\limits_{i=1}^{n}I_{i2}\xrightarrow{p}0$。再由 (12) 式知 $n^{-\frac{1}{2}}\sum\limits_{i=1}^{n}J_{i1}\xrightarrow{\mathcal{L}}N(0,B)$。

结合 $E\{X_i\epsilon_i\}=0$ 和 $E\{[Z_i-\mu(U_i)^TX_i]X_i^T\}=0$ 以及引理4，用类似证明(12)式的方法可以证明 $n^{-\frac{1}{2}}\sum_{i=1}^{n}J_{iv}\stackrel{P}{\longrightarrow}0,\ v=2,3$。另外，由引理4得

$$\Big\|n^{-\frac{1}{2}}\sum_{i=1}^{n}J_{i4}\Big\|\le O_p(\sqrt{n}C_n^2)=o_p(1).$$

这就证明了本引理。

引理6 设条件 $C1$-$C5$ 成立，如果 β 是参数真值，则

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i(\beta)\hat{\eta}_i(\beta)^T\stackrel{P}{\longrightarrow}B.$$

证明 我们仍使用引理5的记号，令 $J_i^*\equiv J_{i2}+J_{i3}+J_{i4}$，那么

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i(\beta)\hat{\eta}_i(\beta)^T &= \frac{1}{n}\sum_{i=1}^{n}J_{i1}J_{i1}^T+\frac{1}{n}\sum_{i=1}^{n}J_{i1}J_i^{*T}+\frac{1}{n}\sum_{i=1}^{n}J_i^*J_{i1}^T+\frac{1}{n}\sum_{i=1}^{n}J_i^*J_i^{*T}\\ &\equiv A_1+A_2+A_3+A_4.\end{aligned}$$

结合(12)式知

$$\begin{aligned}A_1 &= \frac{1}{n}\sum_{i=1}^{n}I_{i1}I_{i1}^T+\frac{1}{n}\sum_{i=1}^{n}I_{i1}I_{i2}^T+\frac{1}{n}\sum_{i=1}^{n}I_{i2}I_{i1}^T+\frac{1}{n}\sum_{i=1}^{n}I_{i2}I_{i2}^T+o_p(1)\\ &\equiv A_{11}+A_{12}+A_{13}+A_{14}+o_p(1).\end{aligned}$$

由大数定律可得 $A_{11}\stackrel{P}{\longrightarrow}B$。下面证明 $A_{12}\stackrel{P}{\longrightarrow}0$。设 $A_{12,rs}$ 是 A_{12} 的第 (r,s) 个元素，I_{ijr} 是 I_{ij}，$j=1,2$ 的第 r 个分量。利用 Cauchy-Schwarz 不等式可得

$$|A_{12,rs}|\le\Big(\frac{1}{n}\sum_{i=1}^{n}I_{i1r}^2\Big)^{\frac{1}{2}}\Big(\frac{1}{n}\sum_{i=1}^{n}I_{i2s}^2\Big)^{\frac{1}{2}}. \tag{13}$$

由引理5知

$$n^{-1}\sum_{i=1}^{n}I_{i1r}^2=O_p(1),\quad n^{-1}\sum_{i=1}^{n}I_{i2s}^2=o_p(1).$$

因此，由(13)式可以证明 $A_{12}\stackrel{P}{\longrightarrow}0$。类似地可以证明 $A_{1v}\stackrel{P}{\longrightarrow}0,\ v=3,4$。因此，$A_1\stackrel{P}{\longrightarrow}\Sigma$。利用类似的证明方法，可以得到 $A_v\stackrel{P}{\longrightarrow}0,\ v=3,4$，这就证完了本引理。

定理1的证明 结合引理5，经简单计算可得

$$\max_{1\le i\le n}\|\hat{\eta}_i(\beta)\|=o_p(n^{\frac{1}{2}}).$$

再结合引理5，利用与 Qin 和 Lawless[14] 类似的证法可知 $\|\lambda\|=O_p(n^{-\frac{1}{2}})$。因此，对(7)式利用 Taylor 展开可得

$$\hat{R}(\beta)=2\sum_{i=1}^{n}\{\lambda^T\hat{\eta}_i(\beta)-[\lambda^T\hat{\eta}_i(\beta)]^2/2\}+o_p(1). \tag{14}$$

再对 (8) 式利用 Taylor 展开，并经简单计算可得

$$\sum_{i=1}^{n}\left[\lambda^T\hat{\eta}_i(\beta)\right]^2=\sum_{i=1}^{n}\lambda^T\hat{\eta}_i(\beta)+o_p(1), \tag{15}$$

$$\lambda=\Big\{\sum_{i=1}^{n}\hat{\eta}_i(\beta)\hat{\eta}_i^T(\beta)\Big\}^{-1}\sum_{i=1}^{n}\hat{\eta}_i(\beta)+o_p(n^{-\frac{1}{2}}). \tag{16}$$

由 (14)-(16) 式可得

$$\hat{R}(\beta)=\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_i(\beta)\Big\}^T\hat{B}^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_i(\beta)\Big\},$$

其中

$$\hat{B}=n^{-1}\sum_{i=1}^{n}\hat{\eta}_i(\beta)\hat{\eta}_i^T(\beta).$$

再利用引理 5 和引理 6 就证完了定理 1。

定理 2 的证明 结合 (9) 式，利用文献 [4] 中定理 2 类似的证法可得

$$\hat{\beta}-\beta=\hat{\Gamma}^{-1}n^{-1}\sum_{i=1}^{n}\hat{\eta}_i(\beta)+o_p(n^{-\frac{1}{2}}). \tag{17}$$

利用引理6的证法可得 $\hat{\Gamma}\xrightarrow{p}\Gamma$，这里 $\Gamma=E(ZZ^T)-E\{\Phi^T(U)\Psi^{-1}(U)\Phi(U)\}$。结合引理 5 及 Slutsky 定理，由 (17) 式可得

$$\sqrt{n}(\hat{\beta}-\beta)\xrightarrow{\mathcal{L}}N\big(0,\,\Gamma^{-1}B\Gamma^{-1}\big).$$

再由简单计算可知 $\Sigma=\Gamma^{-1}B\Gamma^{-1}$，这就证明了本定理。

参考文献：

[1] Yates F. The analysis of replicated experiments when the field results are imcomplete[J]. Emp J Exp Agric, 1933, 1: 129-142

[2] Wang Q H, Rao J N K. Empirical likelihood for linear regression models under imputation for missing responses[J]. The Canadian Journal of Statistics, 2001, 29: 597-608

[3] 田萍，薛留根. 纵向数据半参数回归模型估计的强相合性[J]. 工程数学学报，2006, 23(2): 369-372
Tian P, Xue L G. Strong consistency of estimators in semiparametric regression model for longitudinal data[J]. Chinese Journal of Engineering Mathematics, 2006, 23(2): 369-372

[4] 薛留根，朱力行. 纵向数据下部分线性模型的经验似然推断[J]. 中国科学 A 辑，2007, 37(1): 31-44
Xue L G, Zhu L X. Empirical likelihood-based inference in a partially linear model for longitudinal data[J]. Science in China Series A: Mathematics, 2007, 37(1): 31-44

[5] 魏传华，吴喜之. 部分线性模型非参数部分的多项式关系检验[J]. 工程数学学报，2008, 25(5): 854-866
Wei C H, Wu X Z. Testing polynomial relationships of the nonparametric component in partially linear models[J]. Chinese Journal of Engineering Mathematics, 2008, 25(5): 854-866

[6] Wang Q H, Sun Z H. Estimation in partially linear models with missing responses at random[J]. Journal of Multivariate Analysis, 2007, 98: 1470-1493

[7] 王宁，张应剑. 基于系数估计的趋势性分析检验变系数模型中的不变系数[J]. 工程数学学报，2008, 25(4): 623-633
Wang N, Zhang Y J. Testing for nonvarying-coefficients in varying-coefficient models based on the trend analysis of the coefficient estimates[J]. Chinese Journal of Engineering Mathematics, 2008, 25(4): 623-633

[8] Xue L G, Zhu L X. Empirical likelihood for a varying coefficient model with longitudinal data[J]. Journal of the American Statistical Association, 2007, 102: 642-654

[9] Fan J Q, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models[J]. Bernoulli, 2005, 1: 1031-1057

[10] You J H, Zhou Y. Empirical likelihood for semiparametric varying-coefficient partially linear regression models[J]. Statistics & Probability Letters, 2006, 76: 412-422

[11] 罗羡华，李元，周勇等. 基于纵向数据的半参数变系数部分线性回归模型[J]. 应用数学学报，2007, 30(3): 540-554
Luo X H, Li Y, Zhou Y, *et al.* Semiparametric varying-coefficient partially linear models with longitudinal data[J]. Acta Mathematicae Applicatae Sinica, 2007, 30(3): 540-554

[12] Wang Q H, Rao J N K. Empirical likelihood-based inference under imputation with missing response[J]. Ann Statist, 2002, 30: 896-924

[13] Wang Q H, *et al.* Semiparametric regression analysis with missing response at random[J]. J Amer Statist Assoc, 2004, 99: 334-345

[14] Qin J, Lawless J. Empirical likelihood and general estimating equations[J]. The Annals of Statistics, 1994, 22: 300-325

[15] Chamberlain G. Efficient bounds for semiparametric regression[J]. Econometrika, 1992, 60: 567-596

[16] Mack Y P, Silverman B W. Weak and strong uniform consistency of kernel regression estimates[J]. Z Wahrsch Verw Gebiete, 1982, 61: 405-415

Empirical Likelihood Inferences for Semiparametric Varying-coefficient Partially Linear Models with Missing Responses at Random

ZHAO Pei-xin[1], XUE Liu-gen[2]

(1- Department of Mathematics, Hechi University, Yizhou, Guangxi 546300;
2- College of Applied Sciences, Beijing University of Technology, Beijing 100124)

Abstract: In this paper, we consider estimations of the semiparametric varying-coefficient partially linear models with missing responses at random. An adjusted empirical log-likelihood ratio function for parametric components is proposed. The Wilks' phenomena is proved and the confidence regions for parametric components are constructed. The optimal convergence rate and the semiparametric efficiency bound of maximum empirical likelihood estimators are derived. Simulation results show that the adjusted empirical likelihood method outperforms the unadjusted empirical likelihood method.

Keywords: varying-coefficient partially linear models; empirical likelihood; confidence regions; missing data

Received: 29 Oct 2008. **Accepted:** 01 Apr 2009.

Foundation item: The National Natural Science Foundation of China (10871013); the Natural Science Foundation of Beijing (1102008); the Natural Science Foundation of Guangxi (2010GXNSFB013051); the Graduate Student Foundation of Hechi University (2008QS-N014).